Overview of Architecture Workspace

Robert J. Robbins, Fred Hutchinson Cancer Research Center

Strategic Planning – Strategic Level - Working Group

Face to Face meeting

November 7-8, 2004

NATIONAL CANCER INSTITUTE

caBIG cancer Biomedical Informatics Grid

Architecture Workspace at a Glance - Composition

Workspace Lead: Arumani Manisundaram NCICB Facilitator: Peter Covitz

Architecture Workspace Membership

- 95 on email alias
- 35 Attended F2F in Columbus, OH – July 2004
- 80 Attended Joint VCDE/Architecture F2F in Chicago, IL – Oct 2004

Special Interest Groups

- Interface Architecture
- Information Architecture
- System Architecture
- Security and Access Control
- Software Development - Best Practices & Standards
- Identifiers
- Workflow
- Common Query Language

Architecture Workspace at a Glance
- Composition

Funded Participants

- Cold Spring Harbor
- Duke University
- Fox Chase
- Fred Hutchinson
- Georgetown University—Lombardi
- Ohio State University—Arthur G. James/ Richard J. Solove
- University of Chicago
- University of Pittsburgh
- Washington University—Siteman

Architecture Workspace at a Glance - Composition

Volunteer Participants and Affiliated Organizations

- 9Star Research
- City of Hope
- Columbia University
- First Genetic
- GE Global Research Center
- IBM
- Memorial Sloan Kettering
- NCI
- NCICB
- NCI—CCR
- Oregon Health and Science University
- University of Iowa—Holden
- University of Wisconsin

Architecture WS Activities to Date

- Formation of the following Special Interest Group (sub-groups)
  - Interface Architecture
  - Information Architecture
  - System Architecture
  - Security and Access Control
  - Software Development – Best Practices and Standards
  - ID Management (In collaboration with Data Provenance and Versioning)
  - Workflow
  - Common Query Language

Architecture WS Activities to Date

- caGRID – Phase I (Prototype)
  - Requirements analysis for caCORE virtualization
  - First iteration use cases for
    - Advertisement
    - Discovery
    - Query
    - Object Mapping
    - Semantic Mapping
  - Evaluation of candidate technologies
  - Implementation of caGRID prototype
  - Web presentation/demonstration to caBIG Architecture Workspace
  - caGRID white paper released
  - caGRID-Phase I Source Code released

Architecture WS Activities to Date

- Face-to-Face Meetings
  - Architecture-July 2004
  - Joint Architecture/VCDE Meeting-October 2004
- caBIG Architecture Workspace Mentoring Team Member Assignment
  - Architecture WS participants to provide assistance/guidance to developers in Domain WS to reach Silver level of Compatibility

Key Decisions at Face-to-Face # 1 (Columbus, OH – July 2004)

- Object-oriented Information models will form the basis for expressing data on the grid
- Information exchange on the grid will be represented in XML, and each XML document/message must be validated by an XML Schema
- Models and associated data elements will be registered as metadata in the caDSR
- Classes, attributes and, where possible, data on the grid should be mapped to ontologies. Controlled vocabularies should be used for naming these components. These mappings must be captured in the registered metadata
- The open source Globus and OGSA-DAI toolkits, the caGRID technology extensions, and the Ohio State Mobius technology will form the basis for the next round of data grid prototyping activities.
- Agreed on need for a solution to universal stable identifier for data objects
- Use Enterprise Architect as the UML Modeling Tool

Key Decisions at Face-to-Face # 2 (Chicago, IL – Nov 2004)

- The caGRID team within the Arch. workspace will embark on Phase II
  - Design & Prototyping based on Domain Workspace Use Cases
  - Particular Domain workspace projects will be use as vehicles for testing implementation strategies
  - These activities will yield
    - caBIG “Gold” Level Specification
    - “HOW-TO” – Guide for instantiating a caBIG Gold Grid Service
- Each data object and grid service must have a resolvable unique identifier.
  - Identifiers Subgroup will see look into the LSID specification to check if it provides the necessary framework for such an identifier system, but some additional specification would be needed
  - Provide a final design recommendation by December
- XML Schema alone may be insufficient to describe the complete semantics of a given data class or service
  - An accompanying metadata representation will be necessary
  - RDF/RDFS was proposed for this purpose
  - The caGRID team will evaluate RDF versus traditional XML and make a final design recommendation by December.

Key Decisions at Face-to-Face # 2 (Chicago, IL – Nov 2004)

- Some level of formal data provenance will be necessary to describe data objects that are produced on the grid
  - Develop a minimal, general set of data elements for provenance that could apply to any type of data source.
  - Implementation will allow for extensions to the minimal set that may be needed for a particular sub-domain or measurement technology.

- A common query language that is standardized across all data grid services is needed.
  - The language will be expressible in XML
  - Should interact naturally with the object-oriented data models that will be queried and returned
  - A query Processor layer would be needed at a given data service provider's site
    - Translate the common query into a local query.
  - Need a set of language extensions, or possibly even a separate language, to enable semantic web, inference engine development, and reasoning across ontology structures. (Next Phase)

caGRID – Phase II (Prototype)

caGRID Team Mission

"Define the caBIG system architecture that satisfies the requirements of the caBIG Community"

caGRID – Phase II (Prototype)

caGRID Team Mission

"Define the caBIG system architecture that satisfies the requirements of the caBIG Community"

potential problem?

Are we really a monolithic community?

caGRID – Phase II (Prototype)

caGRID Team Mission

“Define the caBIG system architecture that satisfies the requirements of the caBIG Community”

potential problem?

Are we really a monolithic community?

Or are we a community of communities?

caGRID – Phase II (Prototype)

Single community,
or community of communities?

Addressing this question offers a major opportunity for interaction between the strategic level and architecture workspaces.

caGRID – Phase II (Prototype) - Proposed High-Level – Time Line

	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Requirements Analysis	🟥	🟥	🟥						
System Architecture		🟥	🟥	🟥					
Reference Implementation	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
Technology Evaluation	🟥	🟥	🟥	🟥	🟥	🟥			

Summary

- caBIG Architecture Workspace Subgroups Formed
- caBIG Compatibility Guidelines Document – Ver 1.0 - June 2004
  - Currently aggregating feedback from Domain and VCDE WS
  - Ver. 2.0 will be separated into two documents:
    - Compatibility Document- addresses Data Standards & Interface
    - Best Practices Document- addresses System Architecture
- caGRID – Phase I – July 2004
  - Source Code - Released
  - Presentation - Released
  - White Paper – Released
- caGRID – Phase II – Started- October 2004
  - Use Case / Scenario Review sessions with Domain WS to establish system architecture for caBIG

Summary(cont'd)

- Face-to-Face Meetings to Date
  - caBIG Architecture Workspace – Face-to-Face # 1 – July 2004
  - caBIG Joint VCDE/Architecture – Face-to-Face # 2 – Nov 2004
- caBIG Architecture Workspace Mentoring Team Member Assignment

Next Steps

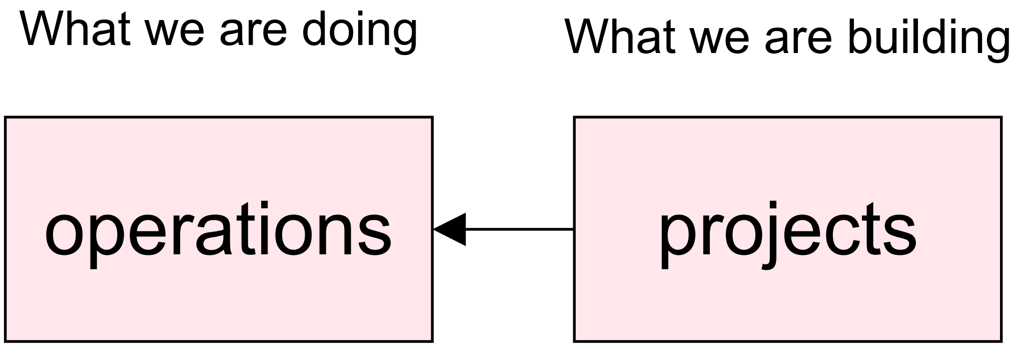
- Identifier subgroup to determine if LSID is sufficient to meet the needs of caBIG
- Workflow subgroup to evaluate workflow technologies and provide a white paper with recommendations for caBIG
- Common Query language subgroup to recommend a common standardised query language across the grid
- Release of Compatibility Document Version 2.0
- caGRID Team will:
  - Design & Prototyping based on Domain Workspace Use Cases
  - Identify particular Domain Workspace projects to use as vehicles for testing implementation strategies
  - Evaluate RDF and XML as the format for metadata representation at runtime
  - Interact closely with Arch and VCDE WS to address all issues being raised in the Workspaces
- caBIG Architecture Workspace Mentoring Team Members to begin interactions with Domain WS Developer Projects

Strategic Planning – Fourth Box Thinking

What we are doing

operations

Strategic Planning – Fourth Box Thinking



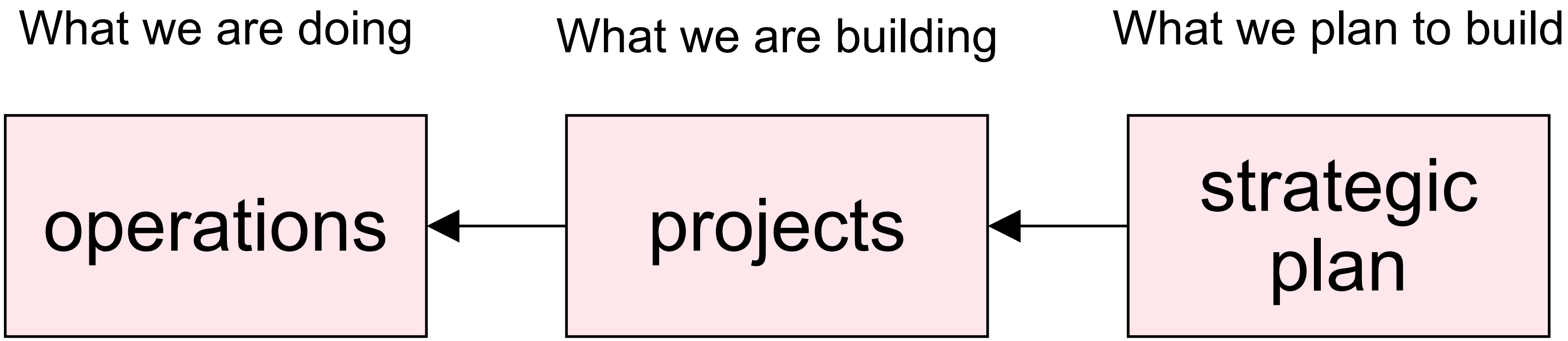
What we are doing

What we are building

operations

projects

Strategic Planning – Fourth Box Thinking



What we are doing

What we are building

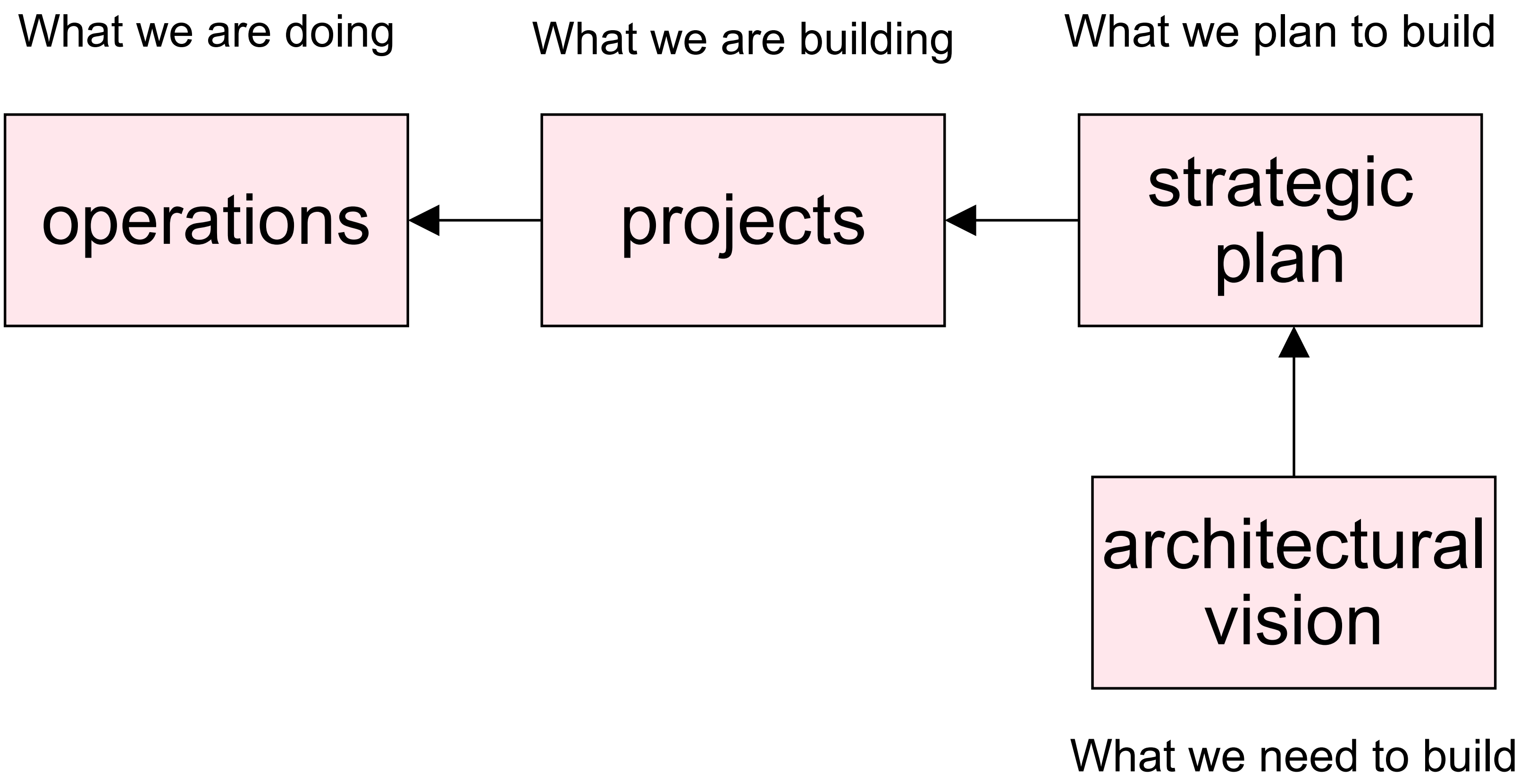
What we plan to build

operations

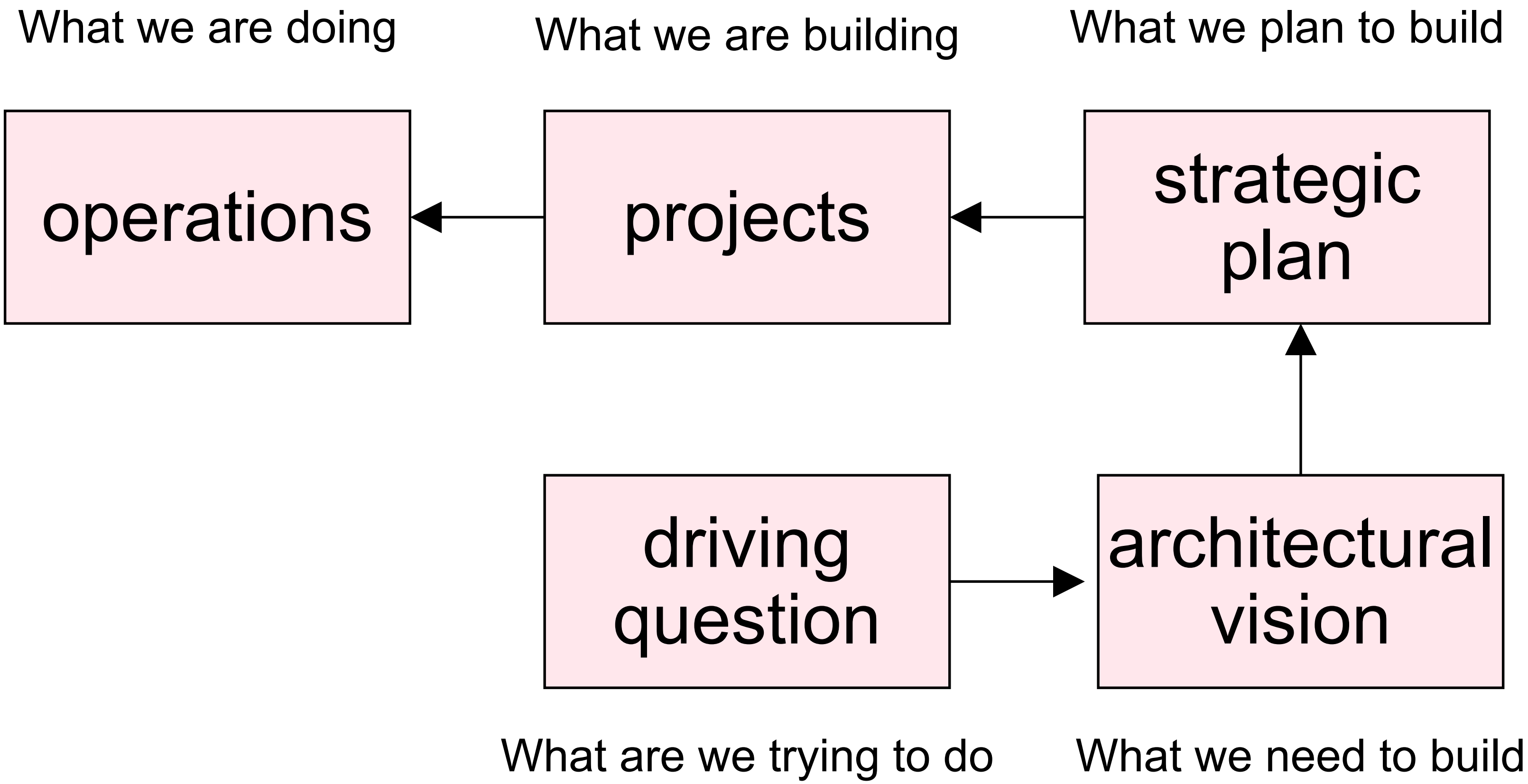
projects

strategic plan

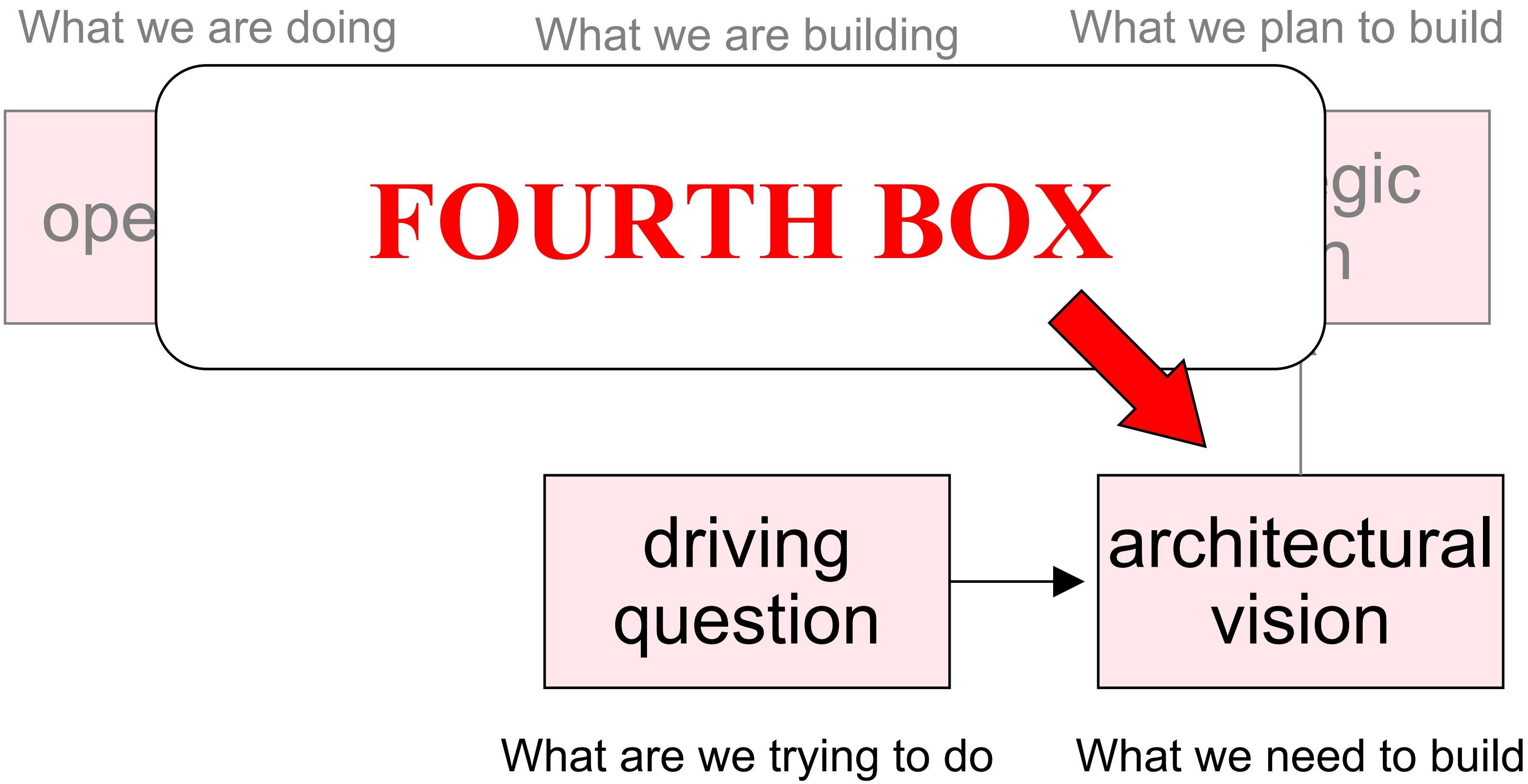
Strategic Planning – Fourth Box Thinking



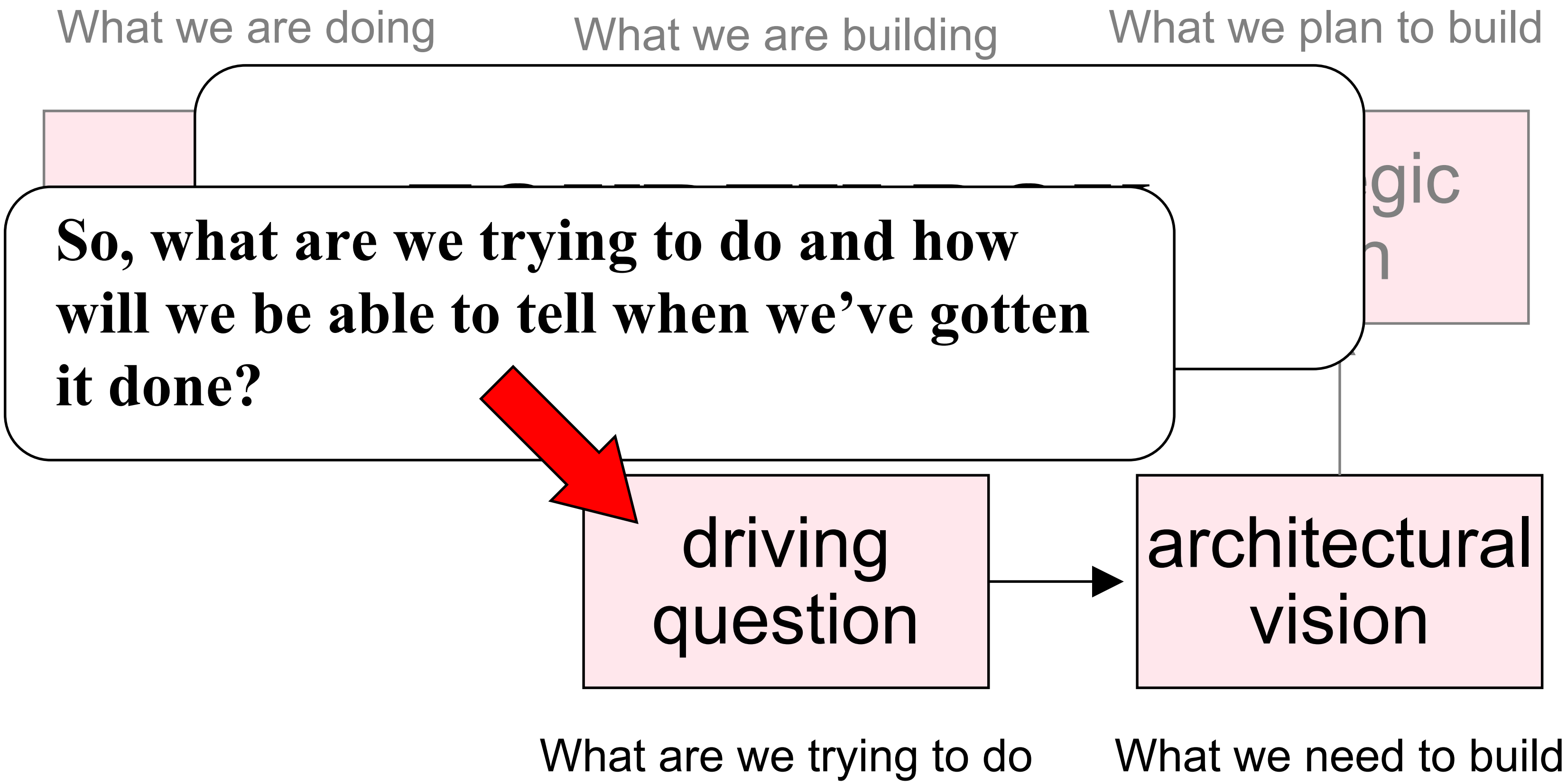
Strategic Planning – Fourth Box Thinking



Strategic Planning – Fourth Box Thinking



Strategic Planning – Fourth Box Thinking


What we are doing

What we are building

What we plan to build

So, what are we trying to do and how will we be able to tell when we've gotten it done?

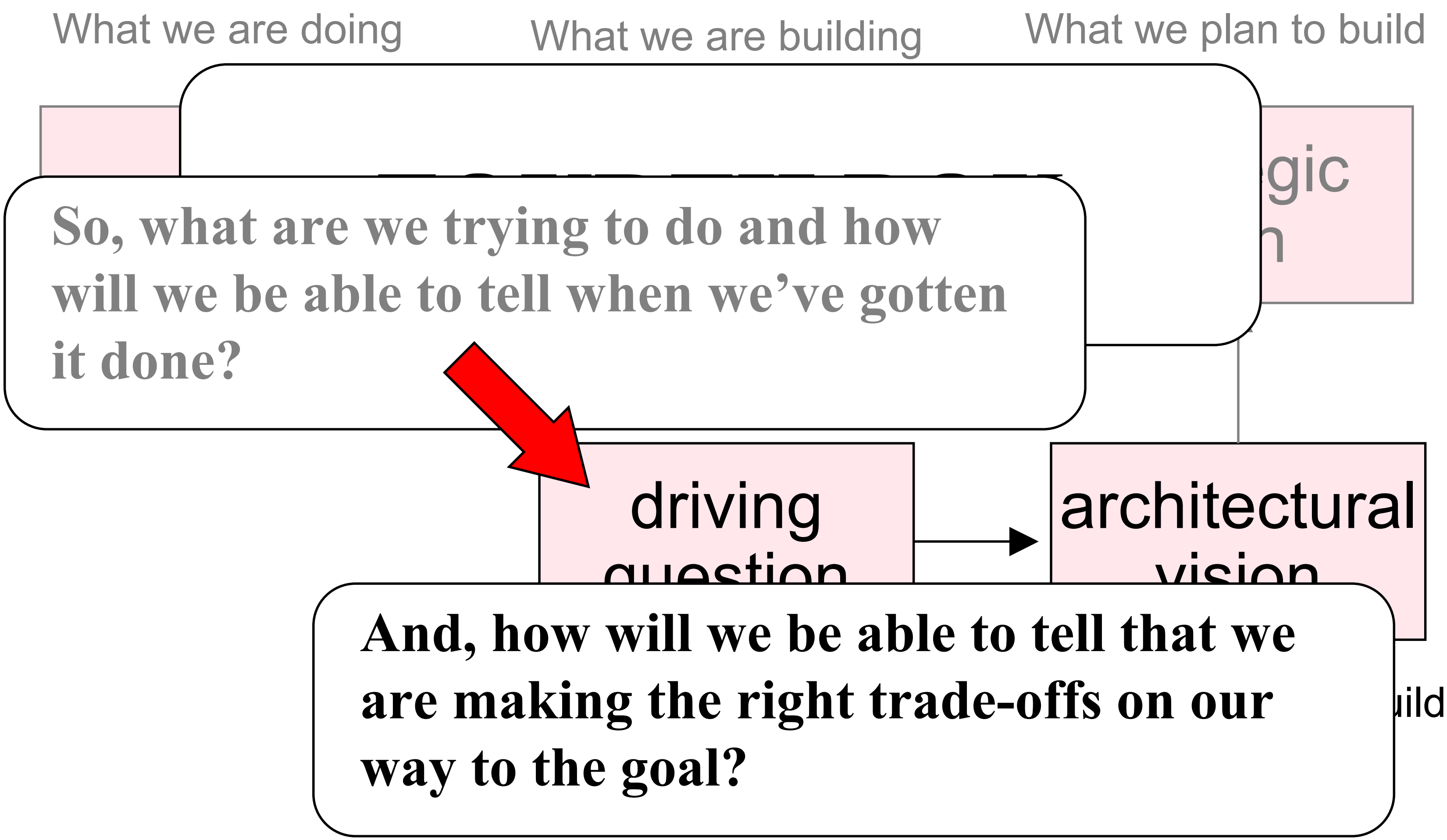
driving question

architectural vision

What are we trying to do

What we need to build

Strategic Planning – Fourth Box Thinking



So, what are we trying to do and how will we be able to tell when we've gotten it done?

And, how will we be able to tell that we are making the right trade-offs on our way to the goal?

Strategic Planning – Fourth Box Thinking

Example of important driving question:

Q: How could you design a communication system that will continue to function, even when pieces have been totally destroyed?

ld

way to the goal?

Strategic Planning – Fourth Box Thinking

Example of important driving question:

Q: How could you design a communication system that will continue to function, even when pieces have been totally destroyed?

A: ARPANET packet-switched network

ld

way to the goal?

Strategic Planning – Fourth Box Thinking

Example of important driving question:

Q: How can you get different networks, using different computers and different operating systems and different network protocols to interoperate?

ild

way to the goal?

Strategic Planning – Fourth Box Thinking

Example of important driving question:

Q: How can you get different networks, using different computers and different operating systems and different network protocols to interoperate?

A: TCP / IP (the INTERNET)

ld

way to the goal?

Strategic Planning – Fourth Box Thinking

Example of important driving question:

Q: How could you separate business logic from the technical manipulation of the contents of databases?

ld

way to the goal?

Strategic Planning – Fourth Box Thinking

Example of important driving question:

Q: How could you separate business logic from the technical manipulation of the contents of databases?

A: The RELATIONAL MODEL of databases.

ıld

way to the goal?

Strategic Planning – Good Solutions

- Highly abstracted components
- Layered architecture
- Modular construction
- Clearly defined interfaces
- No interactions except through interfaces
- Declarative user interface

Strategic Planning – Getting the Right Primitives

Relational database:

Packet-switched networks:

Community information infrastructure:

Strategic Planning – Getting the Right Primitives

Relational database:

Start with semantic-free syntax, add meaning in the upper layers.

Packet-switched networks:

Community information infrastructure:

Strategic Planning – Getting the Right Primitives

Relational database:

Start with semantic-free syntax, add meaning in the upper layers.

Packet-switched networks:

Start with no guarantee of packet delivery, add reliable file transfer in the upper layers.

Community information infrastructure:

Strategic Planning – Getting the Right Primitives

Relational database:

Start with semantic-free syntax, add meaning in the upper layers.

Packet-switched networks:

Start with no guarantee of packet delivery, add reliable file transfer in the upper layers.

Community information infrastructure:

Start with the assumption of total community disagreement, build consensus later.

Strategic Planning – Dueling Aphorisms

We have to build something real now, theory is not enough.

Strategic Planning – Dueling Aphorisms

We have to build something real now, theory is not enough.

If you don’t have time to do it right the first time, how will you ever find the time to do it right the second time?

Strategic Planning – Dueling Aphorisms

We have to build something real now, theory is not enough.

If you don’t have time to do it right the first time, how will you ever find the time to do it right the second time?

We don’t have to wait for perfection in the architecture space before moving forward.

Strategic Planning – Dueling Aphorisms

We have to build something real now, theory is not enough.

If you don’t have time to do it right the first time, how will you ever find the time to do it right the second time?

We don’t have to wait for perfection in the architecture space before moving forward.

True, but we should wait for adequacy. Remember, “lights better” solutions are initially attractive, but ultimately fruitless.

Remember

Success = Deliverables / Expectations

Comments / Discussion?

?

Road to Perdition

Just about every fundamentally flawed IT project began with someone who said,

"Who would ever need more than…"